Unit 3: Opinions and polarization

READINGS AND RESOURCES

- Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, *3*, 87-98.
- Flache A, et al. (2017) Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation* 20(4): 2.

Opinions, attitudes, and beliefs

The thing about the contagion model we studied in the last unit is: contagion is binary. You're either infected or you're not. You either adopt a product or a behavior or a belief, or you don't. This is a good model for some situations, but not for others. In particular, if we're interested in how social information spreads, we have to get a bit more nuanced.

Opinions, beliefs, and attitudes are all somewhat different from one another, but for the sake of tractability here, I'll treat them equivalently and stick with the term *opinion.* An opinion here is a position on some issue. Do you like cake? Is Batman really a hero? Is the theory of natural selection true? In addition to being ruthlessly for or against a particular opinion, individuals can also be ambivalent, or lean weakly for or against. Moreover, the opinion itself doesn't have to be a binary yes/no question. They can be quantitative. How much should one pray? How many licks does it take to get to the Tootsie Roll center of a Tootsie Pop?

In this unit, we'll use models to explore how our assumptions about social influence affects the long-run distribution of opinions in a population.

Building a model of social influence

A model of social influence requires assumptions of three categories.

1. A representation of opinions, attitudes, or beliefs.
2. A mechanism for social influence
3. A population structure.

Let's look at these in more detail.

<u>A representation of opinions, attitudes, or beliefs.</u>
Ultimately, opinions might be complex enough to warrant a sophisticated cognitive model of their representation, such as a semantic network. However, we want to start simple. Let's assume an individual can hold an opinion on any number of topics. We can therefore represent all their opinions as a vector in which each item is one opinion. The values an opinion

can take might be discrete or continuous. If we consider only one opinion with only two possible values, we arrive back at our contagion model from the previous unit, in which opinions are represented as either held or not. In the interest of sticking with simple models and adding only a little bit of complexity at a time, we'll continue in this unit by sticking with agents who hold only one opinion, but now that opinion can take on any continuous value between two extremes. We'll represent an opinion on whatever issue as a real number between 0 and 1, with those numbers representing the most oppositional and extreme opinions an individual can hold, and values in between representing moderate or ambivalent opinions.

<u>A mechanism for social influence</u>
If two agents interact, how do they influence one another? How do opinions change as a result of that influence? In our contagion model, each simply had a fixed probability of "infecting" another agent with its belief. That won't work in the case of continuous-valued opinions. The psychological literature has produced a large amount of research on social influence, and we can formalize some of those ideas here.

- *Positive influence*. When agents interact, they will share their opinions and become more similar to one another.
- *Bounded confidence*, also called biased assimilation. The idea is that individuals are readily influenced by those with whom they are already in substantial agreement, but are more skeptical of opinions shared by individuals with whom they differ. We can represent this with a threshold for influence – if agents are sufficiently different in opinion, they do not influence one another.
- *Negative influence.* An alternative view, for which there is more mixed empirical support, is one in which individuals who differ actually grow more dissimilar. In other words, interactions with those whose opinions differ substantially from one's own can push people toward more extreme versions of their initial views.

We will explore versions of the model with both bounded confidence and negative influence.

<u>A population structure</u>
Who interacts with whom? In the previous unit, we saw that the rate of population mixing could influence the spread of a contagion. Here, we'll focus on two extremes. At one extreme, we'll consider a random mixing models, in which individuals pair up at random. At the other extreme, we'll consider a population with a rigid, unchanging network structure, in which each individual only interacts with four other individuals. This will give us some intuition about how population structure influences opinion dynamics.

The Bounded Confidence Model

CODE: opiniondynamics_BC.nlogo

We'll build a simple opinion dynamics model, based on a 2000 paper by Deffuant et al. We'll make a population of agents each characterized by a single opinion, and for simplicity, we'll initialize those opinions as a random drawn in the range 0 to 1. Each tick of the simulation clock, two agents will interact, and if they are sufficiently similar, they will become more similar. How will this work? Well, let's assume two opinions x_1 and x_2, a confidence threshold d, and a learning rate k. If the absolute difference between x_1 and x_2 is greater than or equal to d, the two agents ignore each other. Otherwise, the opinion of agent 1 is updated like this:

$$x_1 \leftarrow x_1 + k\left(x_2 - x_1\right)$$

In other words, agent 1's opinion moves toward agent 2's opinion, with k representing the proportion of the distance moved. Agent 2's opinion will update similarly toward that of Agent 1.

For the purpose of visualization, an agent's color can represent it's opinion, black for 0, white for 1, and shades of gray in between. How does an agent choose whom to interact with? We will put all the agents on a square lattice, with each agent occupying on cell on the grid. For our initial analyses, we'll ignore spatial location, and simply choose an interaction partner from the population at random. For spatial simulations, we'll restrict interaction partners to the closest four neighbors (up, down, left, or right) on the grid. This is sometimes called the von Neumann neighborhood, after the mathematician and computer science pioneer John von Neumann.

The entire gambit of social network analysis is that interactions are not random, but occur based on social ties. The square lattice is not the most realistic network structure, but it is among the easiest to model computationally, and it gives us some intuition for how structure influences social dynamics. For this model, we will also explore several options for plotting information about a model's dynamics.

SETTING UP THE MODEL

- *learning-rate* slider
- *confidence-threshold* slider
- *spatial-interaction?* switch
- turtles-own [*opinion*]

INITIALIZATION

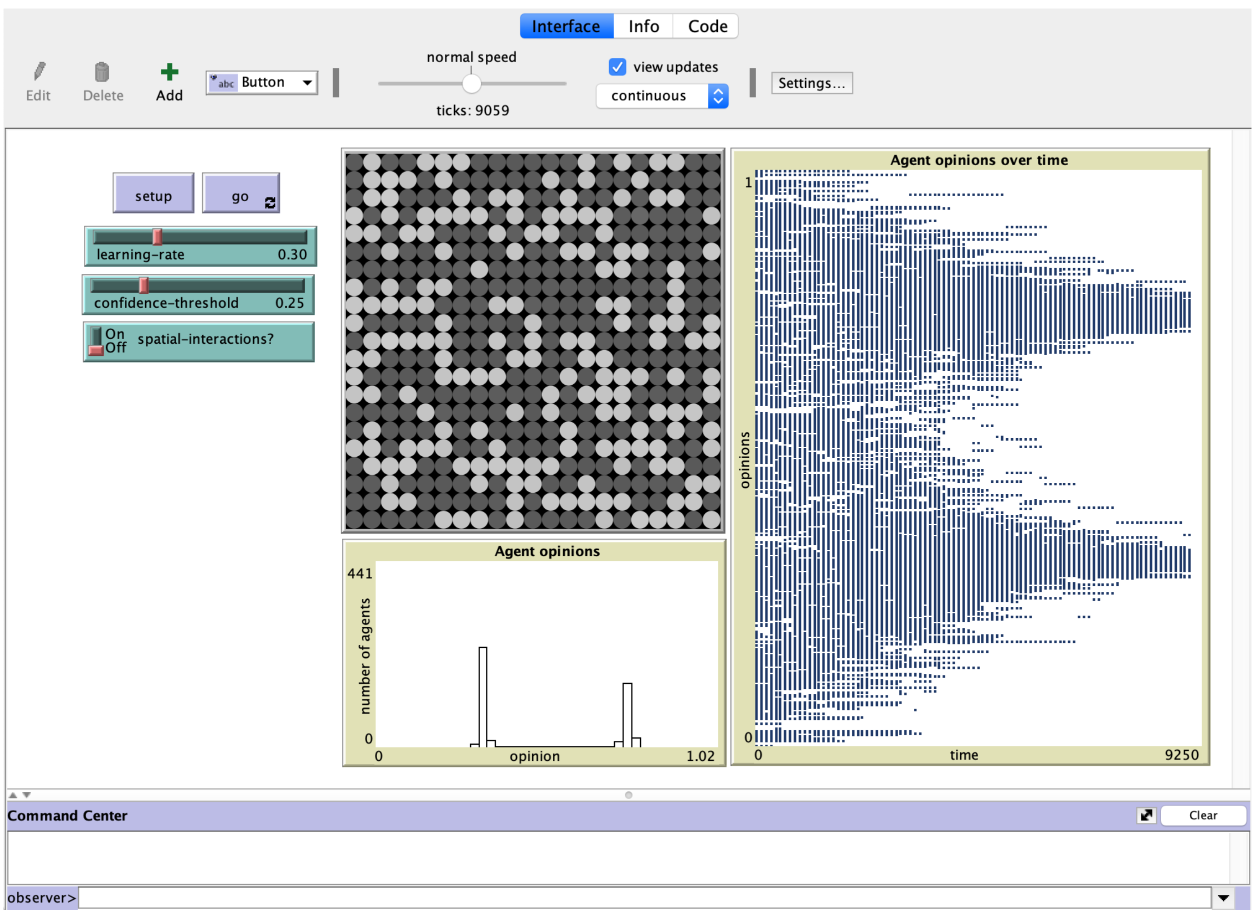
- Each patch of the grid will sprout a turtle. Give each a random opinion in [0, 1].

DYNAMICS

- ASK ONE-OF TURTLES:
  - Final a partner (spatial or random)
  - If they are within the threshold, both turtles update their opinions toward each other.

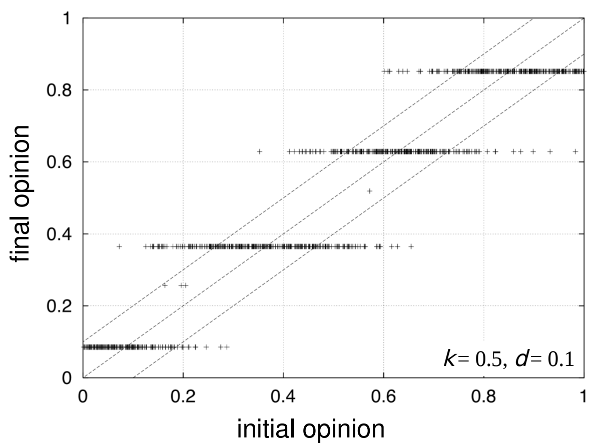
PLOTTING:

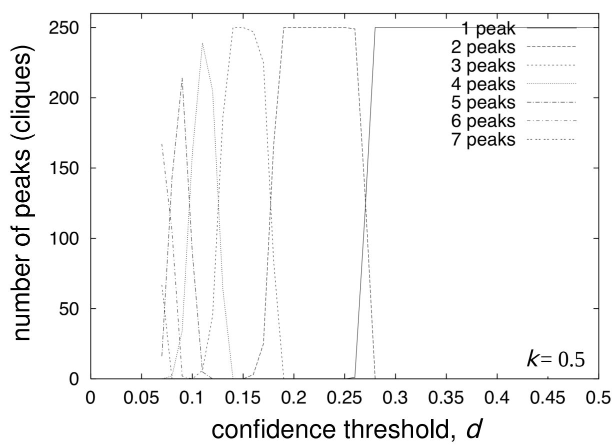
- The agents' colors will represent their opinions, and we can see the change happen on that level.
- Unlike with contagion, there isn't an obvious population-level statistic that summarizes the population. Instead, we are interested in the distribution of opinions. I will visualize this in two ways. The first is as a histogram of opinions. This is useful to represent the distribution, but unless we are watching it very carefully, it doesn't tell us much about the dynamics. The second is to plot the distribution of agent opinions as a temporal scatterplot. This plots opinions as a function of time, and allows us to see how the distribution changes over time, though it can obscure some stuff about density. In combination with the histogram, it's excellent.



RESULTS (non-spatial model)

- If the confidence threshold is large (d = 0.5), we get consensus. This is true even if individuals initially exclude extremely different opinions, as long as the initial distribution of opinions covers the full range of possibilities.
- A little smaller (d = 0.4), and we see some "wings" – a few individuals with extreme values that persist. Notably, we can get consensus even with d considerably smaller than the full range. For example, d = 0.3 usually gives us global consensus.
- As d gets smaller, we see the emergence of cliques. Smaller d means more cliques. For example, 2 or 3 clicks with d = 0.2, 4 or 5 clicks with d = 0.1.
- Agents' initial positions are correlated with their final position, but there is some uncertainty due to stochasticity in who interacts with whom in what order. See below, Figures 3 and 4 from Deffuant et al. (2000).



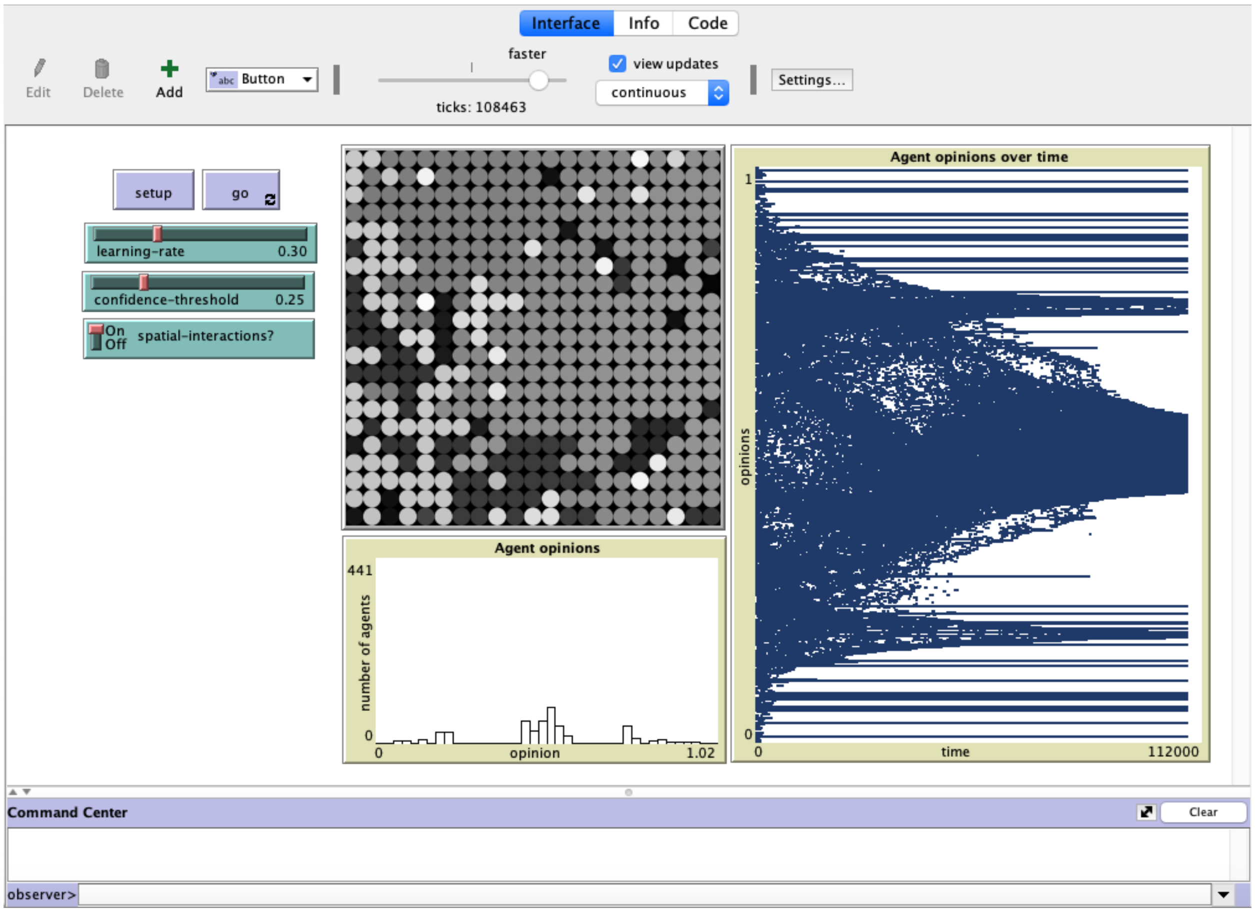


RESULTS (spatial model)

- OK, now let's consider the spatial model, where agents can only sample their immediate neighbors. Even though the square lattice is a very simple sort of social network, comparing it to the well-mixed version can give us an intuition about how the model response to variations in population structure.
- For larger d (d > 0.3), the results are pretty much the same. Mostly consensus. The main difference is the increased likelihood we see a few more "extremists" who don't change.
    - It takes much longer for the population to reach consensus, though, because opinions must percolate along local channels.
    - However, even though we also get consensus, the distribution of opinions around the middle will remain wide for a very long time, because opinions take longer to percolate and the model allows small difference to persist for a long time (because opinion

change is a fixed proportion of the difference in two agents' opinions).

- For smaller *d*, the dynamics can look quite different. Instead of several clusters of the same size as we saw in the non-spatial model, there is often one big cluster that percolates throughout the network and gains a lot of traction. (I got this to happen nicely at d = .25: one big cluster and then a bunch of little spots of other opinions).



- Often, however, much more diversity will instead remain in the population, as many people will have only a few local individuals who are sufficiently similar to them, and opinions cannot therefore easily percolate (spread beyond small local networks).
- So, if we assume strong network structure, we get two seemingly contradictory consequences: most broad consensus in the population among the majority, but more small factions with extreme or idiosyncratic opinions.

Both of these models, the non-spatial and spatial versions, show how group differences can persist even if all influence is positive, as long as differences lead to a lack of influence. However, what if we instead consider the case where differences lead to concrete action: the exacerbation of differences?

Negative influence

CODE: opiniondynamics_neginfluence.nlogo

So far we've assumed that individuals simply ignore others with whom they differ sufficiently. Here, we'll consider the case where those differences are exacerbated. Instead of ignoring those with opinions more than *d* different from one's own, agents will now update their opinions to be increasingly different. Will we still see the emergence of cliques?

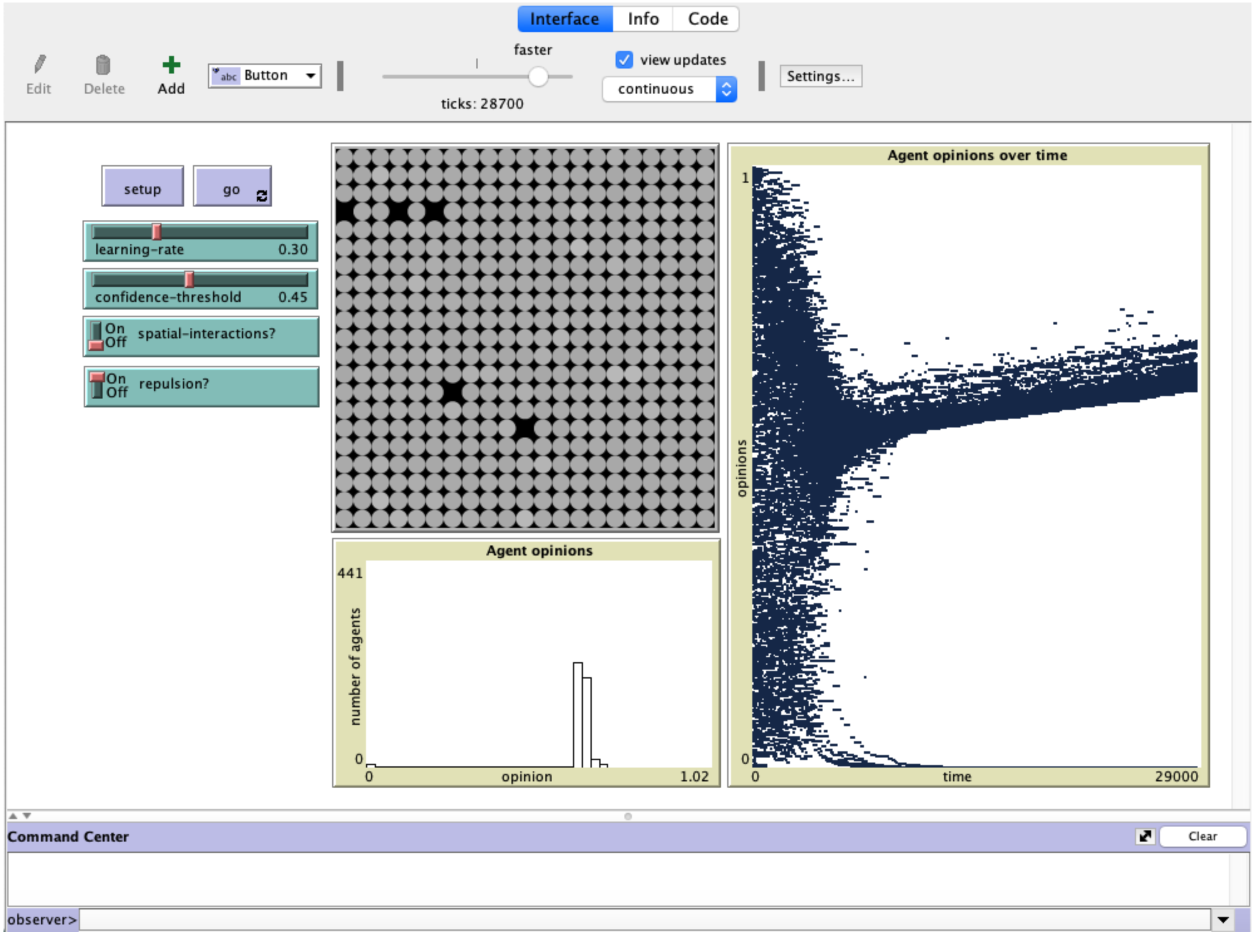
If the difference in agents' opinions are within the confidence threshold, the model works as before, with positive social influence. However, if their opinions are sufficiently different, instead of ignoring each other, they will influence negative influence, become more dissimilar.

CODING THE MODEL
- Add *repulsion?* Boolean switch

INITIALIZATION
- Each patch of the grid will sprout a turtle. Give each a random opinion in [0, 1].

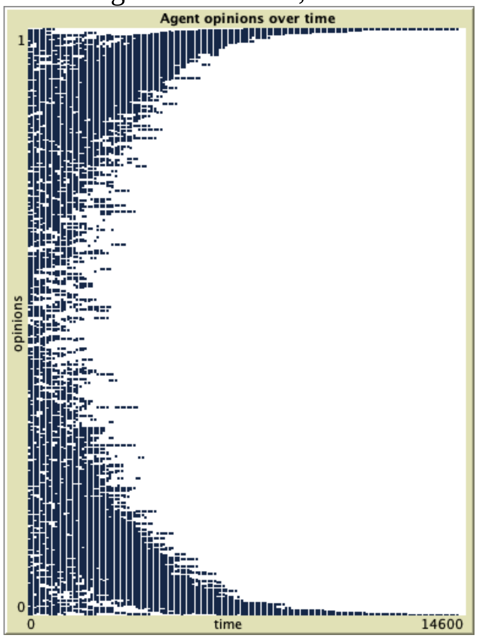
DYNAMICS
- ASK ONE-OF TURTLES:
  - Final a partner (spatial or random)
  - If they are within the threshold, both turtles update their opinions toward each other.
  - **Otherwise, moving the equivalent distance in the direction away from their partner's opinion.**

RESULTS (non-spatial model):
- If *d* is large (*d* > 0.5), the model still leads to consensus.
- When *d* is close to 0.5, we still often get consensus. However, if even a few agents deviate (and become de facto extremists), their existence can now push the rest of the population in the opposite direction. We can see this with *d* = 0.45. This is because the occasional interactions with these extremists put everyone toward the opposite extreme. Thus, the present of even a small number of extremists can push the entire population toward an extreme view if there is sufficiently widespread social influence.

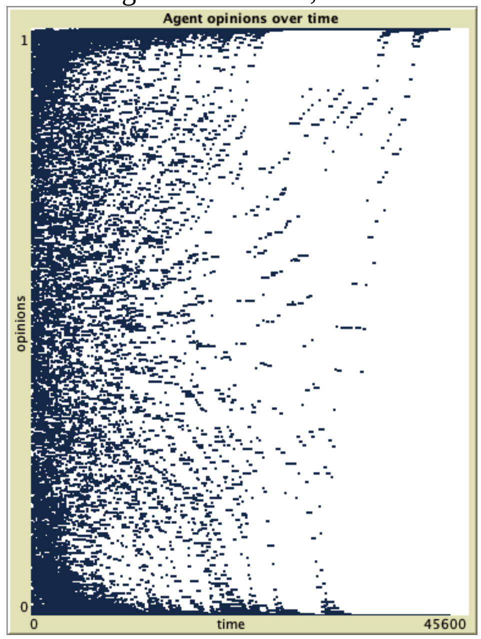


- If *d* gets a bit lower (e.g., *d* = 0.4), the population is instead pushed rapidly toward polarization to the extremes, with roughly equal numbers in both camps.
- When *d* is very low (e.g. *d* = .10), so people are repelled from most others, something interesting happens. You get again attraction to the extremes, but the path there is the perseverance of many intermediate opinions, because those with moderate opinions are constantly pushed in both directions. This will particularly be the case with a small learning rate (e.g., *learning-rate* = .1), so that most opinions changes are small. (check out *learning-rate* = 0.1 and *d* = 0.1).

negative influence, d= 0.4



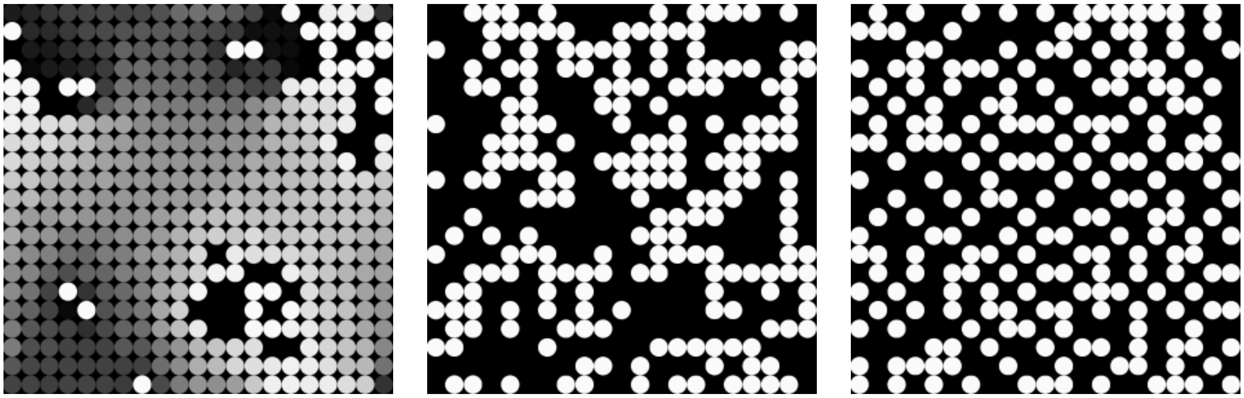
negative influence, d= 0.1



RESULTS (spatial model)

- Now let's consider our spatial model with negative influence.
- If d > 0.5, we get ready clustering around the mean. This looks like consensus, except that there will some persistent variation due to spatial percolation. Any agents near an extremists will be continuously pushed in multiple directions, which maintains the variation.
- When d = .5 and a bit lower, the emergence of polarized zones begins, though it take longer and there is long persistence of variation. The dynamics can take a long time to unfurl, and you get some emergence of demographic zones, including clear boundaries where the views are the most extreme. The closer d is to 0.5, the more continuous the boundaries and wider the territories. (compare d = 0.5 and d = 0.35).
- When d is quite low, you get scattering with a lot of small smatterings of both extremes as everyone is repelled by their neighbors. That is, everyone still holds extreme opinions, but there are no longer blocs of unified opinions. Everyone is isolated.

$d = 0.5$, $t = 70{,}000$ $d = 0.35$, $t = 25{,}000$ $d = 0.1$, $t = 25{,}000$



Among the approaches for modeling social influence and interaction we will consider in this course, opinion dynamics are among the youngest. Fits to empirical data and deep considerations of the cognitive aspects of opinion representation and social influence in the models are still relatively rare. Much of what has been done here is simply to start making some arbitrary assumptions and examine the consequences, many of which aren't particularly intuitive.

I include these models because I think they yield interesting dynamics and provide a good baseline upon which to construct richer models. I also think there is a lot of open territory for the industrious scientist who wants to combine the cognitive science and sociology of opinions. How opinions spread is a very challenging system to model well, because of the inherent difficulties in modeling both complex social interactions and a complex cognitive phenomenon. To do this right, we need to know a lot more about the sociology of communicative interactions and the cognitive science of influence as it interacts with things like language, bias, and identity. The increasing communication between social scientists, cultural evolutionists, and cognitive scientists is encouraging.

Further directions

- <u>Differentiation from similar individuals</u>. Individuals may not simply become more similar to others, but also may actively seek to differentiate themselves from others. Such models have shown that preferences for differentiation can, under a wide range of conditions, nevertheless lead to conformity at the population level. However, if anti-conformists and conformists coexist, anti-conformists can drive the persistence of extreme opinions and therefore drive polarization even in the absence of negative influence.
    - Smaldino PE, Epstein JM (2015) Social conformity despite individual preferences for distinctiveness. *Royal Society Open Science* 2: 140437.

    - Weisbuch G (2015) From anti-conformism to extremism. *Journal of Artificial Societies and Social Simulation* 18(3): 1.
    - Touboul J (2019) The hipster effect: When anti-conformists all look the same. *Discrete & Continuous Dynamical Systems B* 24: 4379–4415.
- Multiple interacting opinions. Individuals have more than one opinion, and this may influence perceptions of similarity and influence, as similarity on sufficient opinions may facilitate positive influence on an opinion on which individuals differ greatly.
    - Axelrod R (1997) The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution* 41: 203-226.
    - Mark NP (2003) Culture and competition: Homophily and distancing explanations for cultural niches. *American Sociological Review* 68: 319-345.
    - Battiston F, Nicosia V, Latora V, San Miguel M (2017) Layered social influence promotes multiculturality in the Axelrod model. *Scientific Reports* 7: 1809.
- Network structure. Here, we considered a very simple network structure (a square lattice). However, real social networks are often more tightly clustered with more varied betweenness among agents. Flache and Macy (2011) considered a vector of opinions and negative influence on a connected caveman graph, and showed how the introduction of long-range ties between clustered could increased the prevalence of extreme positions in the network, providing a possible explanation for how the internet and social media can lead to greater social polarization.
    - Flache, A., & Macy, M. W. (2011). Small worlds and cultural polarization. *Journal of Mathematical Sociology*, 35(1-3), 146-176.
    - Turner, M. A., & Smaldino, P. E. (2018). Paths to polarization: How extreme views, miscommunication, and random chance drive opinion dynamics. *Complexity* 2018: 2740959.
- Scientific beliefs. Some beliefs may be more useful or truer than others. However, it is also clear that misinformation spreads. Some models of social influence have sought to use factors like network structure and conformity to examine these questions, with a focus on scientific communities.
    - Akerlof GA, Michaillat P (2019) Persistence of false paradigms in low-power sciences. *Proceedings of the National Academy of Sciences* 115: 13228-13233.
    - Weatherall JO, O'Connor C (2019) Do as I say, not as I do, or, conformity in scientific networks. *arXiv: 1803.09905v4.*

Exercises

- *Less extreme.* Consider the negative influence model. We assumed that opinions are uniformly distributed in [0, 1]. Instead, assume that this range is narrower, and that opinions are now drawn from [0.5 – X, 0.5 + X], 0 < X ≤ 0.5, so that the baseline model uses X = 0.5. How do smaller values of X affect tendency of the population toward polarization?
- *Ignore them unless they're trouble.* In the negative influence model, we assumed that any agents not being positively influenced were negatively influenced. However, we could also assume a hybrid model in which moderate differences were ignored and only strong differences compelled negative influence. Implement this. How does it affect the resulting opinion dynamics?
- *That's just, like, your opinion, man*. The models of opinion dynamics we have seen have very simple models of individual opinions and beliefs, as well as how those opinions change through social influence. What do you think one or two important questions are about how beliefs and opinions change through social influence? How might a model be constructed to answer that question(s). How would such a model differ from the ones we have seen?